CHAPTER 1

INTRODUCTION

This project uses the Silvaco-SUPREM (Athena-Atlas) as a primary fabrication process and device simulation tool. The first part of this report will discuss the project background underlying the simulation process. Several fabrication processes will be discussed briefly regarding the development of 0.18µm MOSFET and their design issues. The objective and the scope of the project are also mentioned in this chapter.

1.1 Introduction/Project background:

Over the past decades, the MOSFET has continually been scaled down in size, typical MOSFET channel lengths were once several micrometers, but today's integrated circuits are incorporating MOSFETs with channel lengths of about a tenth of a micrometer. Until the late 1990s, the size reduction resulted in great improvement to MOSFET operation with no deleterious consequences. Historically, the difficulties with decreasing the size of the MOSFET have been associated with the semiconductor device fabrication process. For more than 30 years, the integrated circuit (IC) industry has followed a steady path of constantly shrinking device geometries and increasing chip size. This strategy has been driven by the increased performance that smaller devices make possible and the increased functionality that larger chips provide.

Together, these performance and functionality improvements have resulted in a history of new technology generations every two to three years, commonly referred to as “Moore’s Law”. Each new generation has approximately doubled logic circuit density and increased performance by about 40% while quadrupling memory capacity.

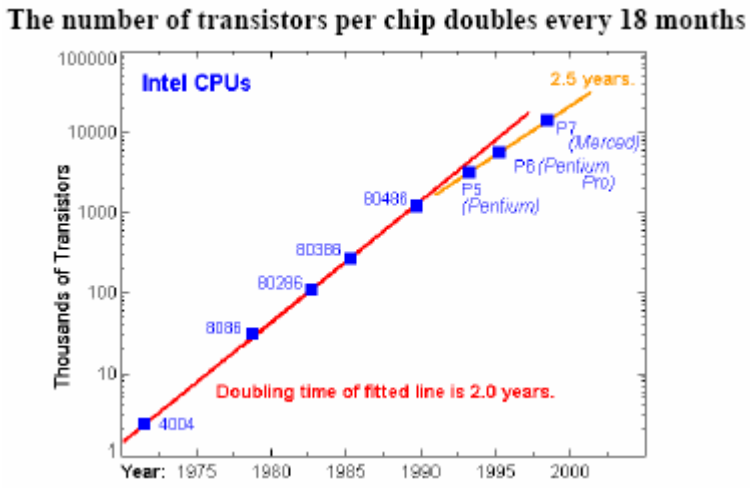

Figure 1.1: Moore’s law (Intel version)

Smaller MOSFETs are desirable for two main reasons. First, smaller MOSFETs allow more current to pass. Conceptually, MOSFETs are like resistors in the on-state, and shorter resistors have less resistance. Second, smaller MOSFETs have smaller gates, and thus lower gate capacitance. These two factors contribute to lower switching times, and thus higher processing speeds. Furthermore, since smaller MOSFETs have lower gate capacitance, and since the amount of charge on a gate is proportional to its capacitance, logic gates incorporating smaller MOSFETs have less charge to move. Indeed, these two factors combined traditionally resulted in a switching times proportional to the squared length of the MOSFET channel. In other words, integrated circuits using 1 micrometre MOSFETs would be roughly 100 times faster than those using 10 micrometre MOSFETs. There is a third reason why MOSFETs have been scaled down in size: smaller MOSFETs can obviously be packed more densely, resulting in either smaller chips or chips with more computing power in the same area. Since the cost of producing integrated circuits is highly related to the number of chips that can be produced per wafer, this third reason for MOSFET scaling is perhaps as important as the first two.

Up until now, MOSFET scaling has proceeded based on the scaling theory without serious roadblocks. MOS transistors with a gate length as short as 10nm, although experimental, have been demonstrated.

The scaling theory, based on a constant electric-field, requires supply voltage, threshold voltage, gate length, and gate oxide thickness to be scaled down by a scaling factor. The doping level in the channel must be scaled up by the same scale factor. The junction depth of source and drain also needs to be scaled down to suppress the short-channel effect. Figure 1.2 shows the cross section of original NMOS transistor and scaled NMOS transistor. Another important aspect of transistor scaling is the scaling of parasitic resistances and capacitances. These parasitic components do not necessarily scale with transistor scaling. Therefore, it becomes increasingly critical to minimize parasitic components in order to get the best return-on-scaling on transistor performance [21]. A good example to address this issue is the silicidation of drain, source, and gate as shown in Figure 1.3. Titanium silicide used in advanced CMOS technology dramatically reduces parasitic resistances in the device.

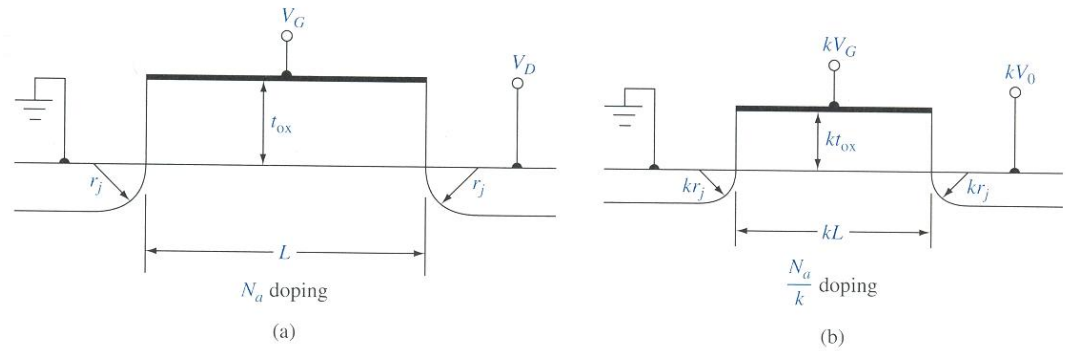

Figure 1.2: Cross section of (a) original NMOS transistor and (b) scaled NMOS transistor.

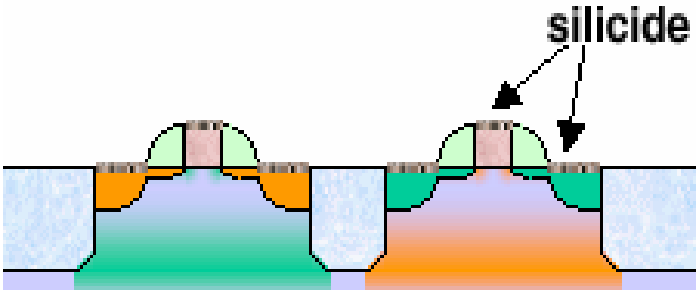

Figure 1.3: Silicidation of drain, source and gate in advanced CMOS technology

Transistor scaling, in practice, has not followed exactly the constant E-field scenario. For performance reasons and due to product requirements, scaling of supply voltage did not happen as fast as geometrical scaling, such as gate length and gate oxide thickness. Because of this, the electric field in the device increased with scaling, resulting in aggravation of short-channel effect (SCE). Short channel effects impact threshold voltage, subthreshold currents, and I-V behavior beyond threshold. In addition, it also increased reliability concerns such as hot carrier effect (HCE) and gate oxide reliability. Various transistor design techniques have been proposed and investigated to deal with SCE and HCE.

One of the most important developments in transistor design to deal with SCE and HCE is the use of lightly doped drain (LDD) in conjunction with polysilicon gate sidewall spacer (Figure 1.4). This technique, introduced in the industry at late 1970's, has become a standard feature for sub-micron transistors, typically having gate lengths of 0.50um and below. Various other ideas have been proposed and adopted in transistor design. These include retrograde channel doping, super-steep retrograde channel, halo or pocket implant with a large tilt angle, pre-amorphization implant (PAI), and source/drain extension [21].

In today's advanced CMOS technology, MOS transistors are typically implemented in a dual gate CMOS configuration: n+ poly gate for NMOS, and p+ poly gate for PMOS (Figure 1.4). Dual-gate CMOS allows both N-channel and P-channel transistors to operate in a surface-channel mode. However, it presents process integration issues such as boron penetration in the p+ poly gate, which causes device instability and gate oxide reliability problems in the P-channel MOSFET. Reduction of the thermal budget to minimize boron penetration can cause

insufficient dopant activation in the gate poly, leading to poly depletion problems. These are some of the challenges in further scaling of MOSFET below a 100nm gate length

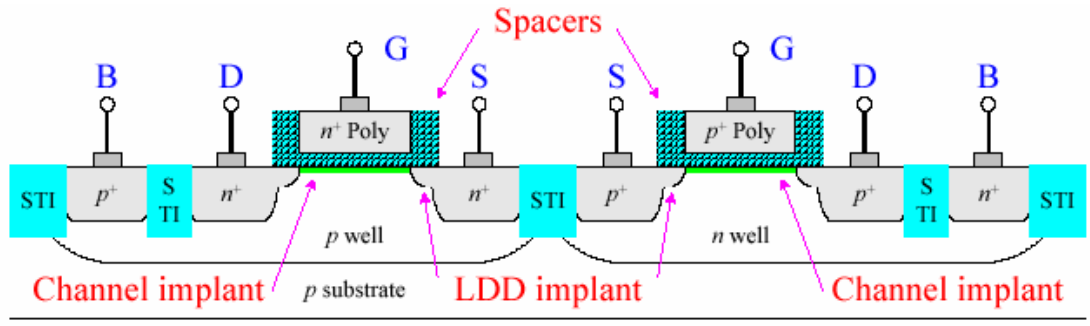

Figure 1.4: Dual gate CMOS configuration: n+ poly gate for PMOS

The project focused on the Metal Oxide Semiconductor Field Effect Transistor (MOSFET), which has been the most important device for today's advance Integrated Circuit (IC) industry. The MOSFET, which has a simple structure as compared to its BJT counterpart, the sizes have shrunk from a few micrometers to less than quarter micrometer. However, continuous shrinking in the device size has caused the conventional one-dimensional MOS transistor theory to be insufficient to explain the deep-submicron MOSFET thoroughly. A 0.18μm MOSFET (NMOS and PMOS) were developed according to the ITRS roadmap (Table 1.1). Figure 1.5 shows the cross section of a MOSFET transistor. The focus of this report will be on discussing the short channel effect whenever a MOSFET channel length is decreased and applying current technology to prevent the short channel effect. Several factors which causes a threshold voltage variation and design analysis were also performed.

Table 1.1: MOS scaling requirements from the ITRS roadmap

Year	1999	2003	2006	2009	2012
Technology node	0.18µm	0.13µm	0.10µm	0.07µm	0.05µm
DRAM Bits/Chip	1G	4G	16G	64G	256G
Minimum Supply Voltage (volts)	1.5 – 1.8	1.2 – 1.5	0.9 – 1.2	0.6 – 0.9	0.5 – 0.6
Gate Oxide τ_{ox} Equivalent (nm)	3 - 4	2 – 3	1.5 - 2	< 1.5	<1
Contact Xj (nm)	70 – 140	50 – 100	40 – 80	15 – 30	10 - 20
Xj at Channel (nm)	36 – 72	26 - 52	20 -40	15 – 30	10 - 20
Minimum Logic V_{dd}	1.5 – 1.8	1.2 – 1.5	0.9 – 1.2	0.6 – 0.9	0.5 – 0.6
Xj at source/drain extension	42 – 70	24 – 40	20 – 33	16 – 26	11 – 19
Channel Dopant Concentration ($at.cm^{-3}$)	2e18	3.3e18	4e18	5e18	14e18

Xj = Junction Depth

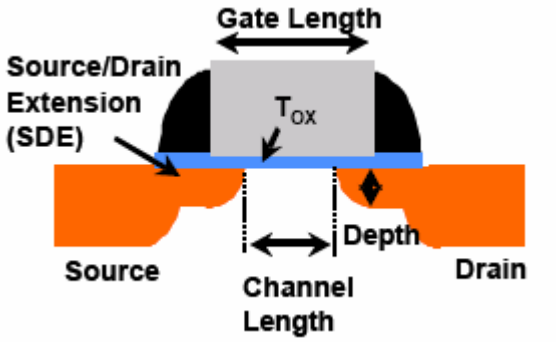

Figure 1.5: Cross-section drawing of a MOSFET transistor

1.2 Objectives

The main objective of the project is to develop a 0.18µm n-channel (NMOS) and p-channel (PMOS) MOSFET according to the ITRS roadmap. Many design aspects has to be considered when the MOSFET device is scaled down into deep submicron regime. Short channel effects that will appear whenever the MOSFET device is scaled down and gate oxide has to be thin

enough to increase the device performance. There were several advanced fabrication processes is applied to the 0.18μm MOSFET design such as halo implant for the punch-through stopper, light doped drain (LDD) to avoid hot electron and retrograde well to suppress the parasitic bipolar devices (latch up immunity). Therefore, the summary of objectives and aims of this project are:

- To apply advance fabrication process to the 0.18μm MOSFET and to study the effects on device performance.
- To study the factors that caused the variation of threshold voltage.
- To study the effectiveness of advanced technique in preventing threshold voltage variation.
- To study the limitations of a MOSFET designs and their solutions.
- To reduce the short channel effects of a deep-submicron device.
- To be exposed to MOSFET design procedures and would be able to reinforce the understanding of MOSFET devices by participating in the device design process.

1.3 Scope and organization

The integration of a new manufacturing process flow has been implemented through the design of 0.18μm MOSFET device. The remainder of this thesis describes the challenges facing device design in the deep sub-micron region, paying special attention to those that have the most relevance for the rest of this thesis: short channel effects. Device design considerations such as channel, halo and retrograde well doping are then presented.

The device technology employed includes: surface channel n+ poly gates for NMOS and buried channel n+ poly gates for PMOS, aluminum metallization, shallow ion implanted sources/drains, twin-well process, punch-through stopper and is designed to operate at a supply voltage of 1.8V. A tolerance analysis was performed using Athena/Atlas simulation in order to develop a robust process that yielded consistent device results. Generally, the project consists of two parts:

1) Process simulation

The process used to fabricate the NMOS and PMOS transistors has been simulated in Silvaco-Athena to verify the correct process parameters such as implant dose and energy, thermal steps, and film deposition, result in the desired doping profiles and device structure.

2) Device simulation

The results of the process simulation program were used as the input for a device simulator (Silvaco-Atlas) and the device characteristics can be examined. This provides an easy way of studying the effects of process parameters on the device performance and both the device structure and the fabrication process can thus be optimized.